AD-A250 505

...ATE	3. REPORT TYPE AND DATES COVERED
	Final 1 Aug 88 31 Jul 91

4. TITLE AND SUBTITLE

Signal Processing and Recognition in Adaptive Neural Networks

5. FUNDING NUMBERS

MD 910318-7388-360201
G--AFOSR-88-0204
PE-61102F
PR-2313 TA-B8

6. AUTHOR(S)

Shihab A. Shamma and P.S. Krishnaprasad

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Systems Research Center , Electrical Engineering Department, University of Maryland, College ark Maryland 97 20742

8. PERFORMING ORGANIZATION REPORT NUMBER

AFOSR-TR 92 0307

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

AFOSR, Attn: Dr. John Tangney
Building 410 BollingAFB, DC 20332-6448

NL

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT

12b. DISTRIBUTION CODE

13. ABSTRACT (*Maximum 200 words*)

The research reported here has been conducted over the last three years under the AFOSR grant (AFOSR-88-0204). It was divided into four general categories of projects: (1) Cochlear models: applications and implementations. Detailed models of the auditory peripheray were developed and applied as front-ends for speech recognition experiments. (2) Early auditory processing: binaural hearing and phonemic segmentation. Physiological plausibility of traditional models was examined and alternative formulations were made and tested. (3) Central auditory function: physiology, psychoacoustics, and mathematical models. Experiments focused on the primary auditory cortex and the anterior auditory field. Models of the responses were applied to generalized representations of speech. Psychoacousticals experiments were carried out to elaborate and test the physiologically derived models. (4) Analysis of neural networks in applications to tactile sensing. Mathematical formulations of the deconvolution problem were analyzed and solved using neural network structures.

92-12882

SUBJECT TERMS	15. NUMBER OF PAGES
	16. PRICE CODE

SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
(U)	(U)	(U)	(U)

SYSTEMS RESEARCH CENTER
and
DEPARTMENT of ELECTRICAL ENGINEERING
University of Maryland
College Park, MD 20742

Final Report

Signal Processing and Recognition in Adaptive Neural Networks

Prepared by

Shihab A. Shamma and P.S. Krishnaprasad

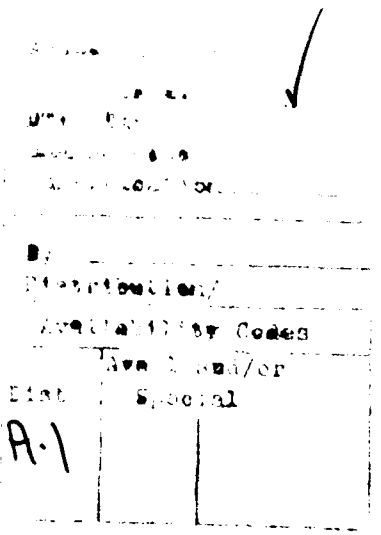
Submitted to

The Air Force Office of Scientific Research
Attention: Dr. John Tangney

AFOSR/NL, Building 410
Bolling AFB, DC 20332-6448

Contact Personnel:

Dr. Shihab A. Shamma (301) 405-6842 (technical)
Ms. Antoinette Lawson (301) 405-6274 (business)



Signal Processing and Recognition in Adaptive Neural Networks

Shihab A. Shamma and P.S. Krishnaprasad

Abstract

The research reported here has been conducted over the last three years under the AFOSR grant (AFOSR-88-0204). It was divided into four general categories of projects: (1) Cochlear models: applications and implementations. Detailed models of the auditory peripheray were developed and applied as front-ends for speech recognition experiments. (2) Early auditory processing: binaural hearing and phonemic segmentation. Physiological plausibility of traditional models was examined and alternative formulations were made and tested. (3) Central auditory function: physiology, psychoacoustics, and mathematical models. Experiments focused on the primary auditory cortex and the anterior auditory field. Models of the responses were applied to generalized representations of speech. Psychoacousticals experiments were carried out to elaborate and test the physiologically derived models. (4) Analysis of neural networks in applications to tactile sensing. Mathematical formulations of the deconvolution problem were analyzed and solved using neural network structures.

Besides the two P.I.s' salaries, the grant supported several Ph.D. and M.S. students, a laboratory manager, and partially a post-doctoral fellow (see list of names and degrees at the end of this review).

I. Cochlear Models: Applications and Implementations

I.1 Applications of cochlear models to speech processing and recognition

Over the last few years, we have carried out experimental and theoretical studies of mammalian auditory processing, particularly with a view towards applying its underlying functional principles in the design and implementation of automatic speech recognition systems. Of direct relevance here are the following two results:

I.1.1 Representation of the acoustic features of speech phonemes in the auditory system

Models of cochlear processing can be used effectively to study the underlying bases of our perception of speech sounds. Specifically, they can answer questions regarding which spectral and temporal features survive the many stages of (nonlinear) processing in the auditory periphery. To address these questions, we carried out extensive experiments in which we first determined the auditory outputs due to various speech phonemes from many speakers (Fig.1b). These outputs are computed using computational algorithms based on detailed biophysical models of the auditory periphery that we developed earlier. The auditory patterns were next applied to single layer feedforward neural networks that were trained to recognize different groups of phonemes according to their *place of articulation* (Fig.1a). By analyzing the connectivities of the resulting trained networks, specific speaker-independent acoustic features can be readily isolated (Fig.1c-d). There are two fundamental findings from these experiments. The first is that the auditory periphery preserves all information necessary to identify speech phonemes. The second is that only a few specific features are important in the recognition of speech phonemes. The experimental paradigms and the theoretical development of the algorithms used in this work are detailed in a M.S. thesis by K. Wang which is appended to this proposal.

Extraction and Enhancement of the Acoustic Features of Phonemes

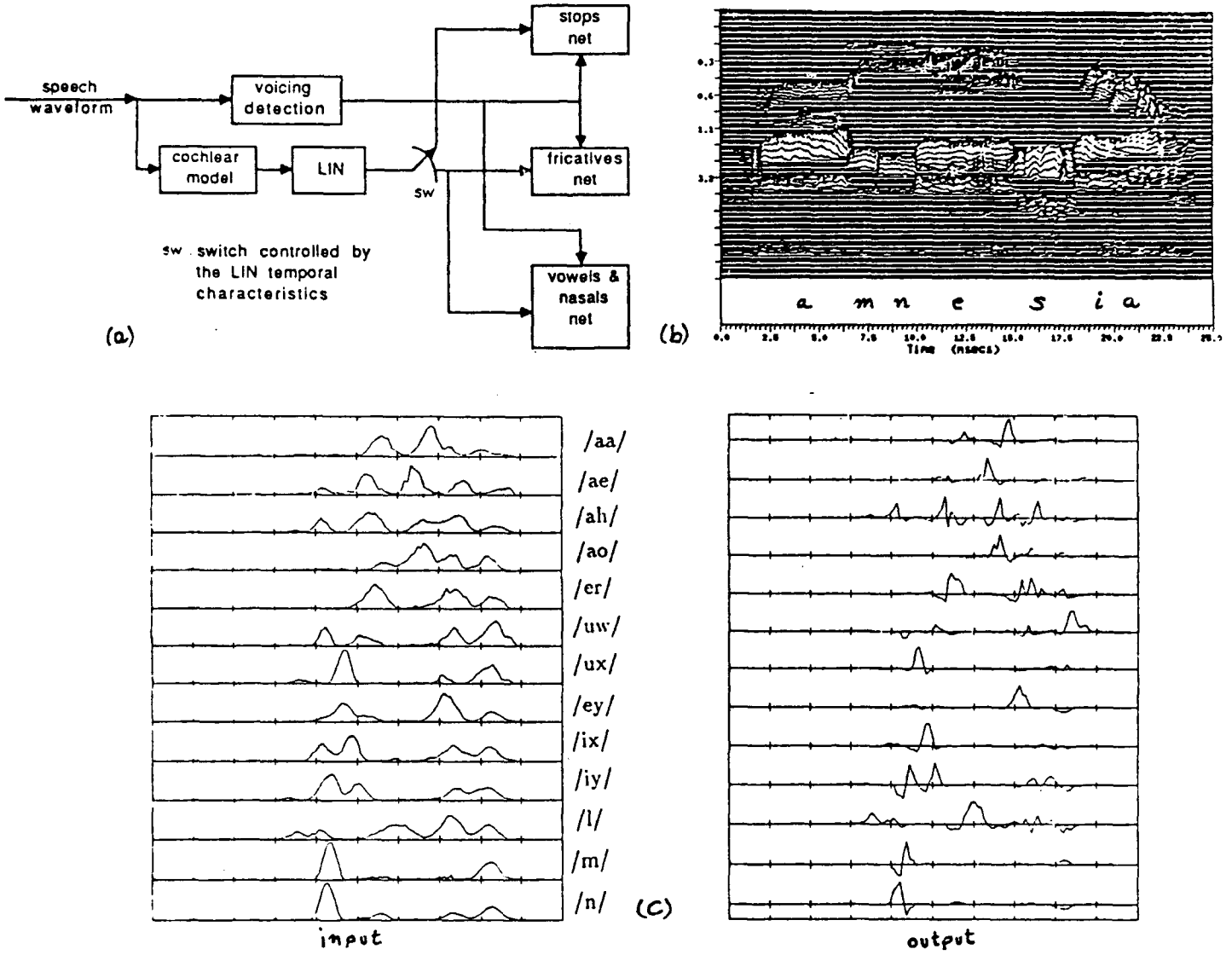


Figure 1: Extraction and enhancement of the acoustic features of speech phonemes. Overall plan of the algorithms used is shown in Fig.1a.. The cochlear model and the lateral inhibitory networks (LIN) are used to generate the auditory patterns shown in Fig.1b. For each class of phonemes, single layer feedforward networks were trained to classify them. Fig.1c illustrates cross sections of the class of vowels and nasals for one speaker. To the left we show the input patterns to the classifier network; To the right is the corresponding features that are most important for the recognition of each pattern.

I.1.2 The performance of cochlear front-ends in noisy speech

We have carried out experiments to assess the performance of cochlear models as front-ends to speech recognition systems under noisy environments. In this work, the outputs of the auditory models are computed for clean speech tokens, and for signals contaminated by broadband noise at various signal-to-noise ratios. The performance of a speech recognizer is then tested for all these outputs, and in turn compared with contaminated speech signals using other parametric and nonparametric representations (e.g., LPC, Cepstral, Spectral). To perform this comparison, common output measures are used such as VQ or KLT vectors. Typical results from four such investigations are illustrated in Fig.2. In each plot, a specific output measure is compared under clean and noisy speech condition for up to four different representations.

For instance, in Fig.2(a) the effect of additive noise on the various representations is measured first as

$$D_{Percent\ \ Distortion} = \frac{1/N \sum_{j=1}^{N} \|K(F(s_j)) - K(F(s_j + n_j))\|_2}{1/N \sum_{j=1}^{N} \|K(F(s_j))\|_2}$$

where $F(s_j)$ is the representation of frame j of the clean speech, N is the number of frames, and $s_j + n_j$ refers to a frame of speech with additive noise.

The Karhunen-Loeve transform, K, is computed for each representation from the autocovariance of the clean speech features. It is chosen as a means of reducing the dimension of the cochlear model in an optimum fashion. Since it also can be used to restrict measured data to a known signal space, it is applied to all representations so as not to give the cochlear model an unfair advantage. The eigenvectors corresponding to the 48 largest eigenvalues of the autocovariance matrix are chosen to form the transform kernel for both the spectral and cochlear representations. For the LPC and LPC cepstrum, all eigenvectors are retained. Using this distortion measure, the cochlear model suffers less distortion than the other representations at noise levels less than 9db, at which point it becomes parallel to the parametric models.

Similar results in general are obtained for the other distortion measures depicted in Fig.2. For instance, the distortion measure in Fig.2(b) is defined as,

$$D_{Distribution\ \ Distortion} = 1 - \frac{\sum_{i=1}^{64} f_s(i) \cdot f_{s+n}(i)}{\sqrt{\sum_{i=1}^{64} f_s^2(i) \sum_{i=1}^{64} f_{s+n}^2(i)}}$$

where f_s and f_{s+n} are the class distribution of the quantized clean speech and the distribution of the quantized noisy speech, respectively, computed using a VQ codebook of 64 symbols. In Fig.2(c), a normalized VQ performance measure is used:

$$D_{VQ\ \ Distortion} = 1/N \sum_{j=1}^{N} \|F(s_j) - VQ(F(s_j + n_j))\|_2.$$

Fig.2(d) illustrates the discrimination ability of the different measures using a variant of the Fischer Discrimination as:

$$D_{Confusion\ \ Score} = 1/n \log \frac{\det S_W}{\det S_B}$$

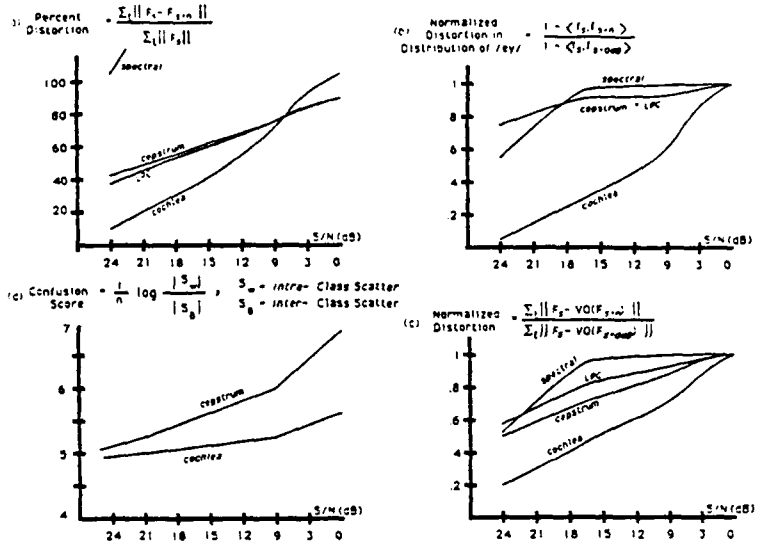


Figure 2: Distortion due to noise: (a) Percent Distortion (b) Change in VQ Class Distribution of phoneme /ey/ (c) Normalized VQ Distortion (d) Intra-Class vs Inter Class Scatter

where $S_W$and S_B are the intra-class and inter-class scatter matrices, respectively

$$S_W = \sum_{i=1}^{c} \sum_{x \in \chi_i} (x - m_i)(x - m_i)^t$$

$$S_B = \sum_{i=1}^{c} n_i (m_i - m)(m_i - m)^t$$

and c is the number of phonemes, χ_i is the collection of all representations, x, labeled as the i^{th} phoneme, n_i is the cardinality of χ_i, and m and m_i are found by averaging all features and averaging all the features in χ_i, respectively.

These results point to a significant advantage for the cochlear front-ends in such tasks. Since then, we have been manipulating the parameters of the cochlear model so as to uncover the origins of its robustness. We have also expanded the experiments to isolated-speech-recognition tasks to ensure that sufficient information is preserved by the cochlear model. In the latter experiments, the cochlear advantage persists at about the same level as illustrated in Fig.3.

Recently, we started adapting the cochlear model to operate in conjunction with the SPHINX system which was supplied to us by Dr. Richard Stern of Carnegie-Melon University. We expect to start experimenting with it within the next two months.

I.2 Implementations of the cochlear and other auditory models

In order for cochlear and other models of auditory processing to become useful computational structures, their computational cost has to be reduced drastically. This is important not only for real-time applications, but also to facilitate experimentation with such models. At present, the fastest implementations of the cochlear models are about a 1000 times slower

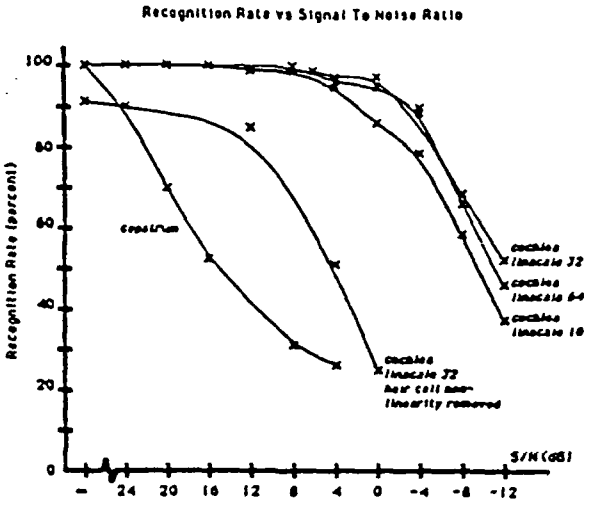


Figure 3: Recognition rate as a function of added noise level using cepstral and cochlear patterns with different parameters. Test set is 16 utterances of the isolated digits 0-9 from the TI data base. VQ codebook used contains 256 symbols for both cochlear and cepstral features.

than real-time. Continued improvements in computational technology is expected to close this gap, but not sufficiently or rapidly enough. The two obvious alternative options are a radical reformulation of the algorithms to minimize their complexity without sacrificing the associated advantages, and the fabrication of specialized hardware.

The first option is only plausible once a clearer understanding of cochlear function is achieved. Several groups have been pursuing the second option, that of fabricating specialized hardware to perform the cochlear operations. These efforts fall into two different approaches: The first aim to map the cochlear algorithms into digital algorithms that are then implemented using standard DSP chips. The second aim to fabricate various digital, analog, or mixed custom integrated circuits. In the first category, only moderate success has been achieved. For instance, we attempted two years ago to implement our cochlear model on the ODYSSEY Board, a TI product which contains four TMS32025 DSP Chips and associated circuitry, and sits on the backplane of the TI Explorer . Three processors worked in parallel to compute the cochlear outputs from 128 channels, while one coordinated all actions and memory transfers. The implementation worked well except for the massive I/O bottlenecks created by the need to pass data back and forth to the computer memory. This slow and unavoidable process offset any advantages gained from speeding up the computations. Overcoming this problem requires specialized parallel interfaces and software which we do not as yet possess.

Fabricating specialized integrated circuits has proven so far equally difficult. Several fundamentally different approaches have been tried, ranging from the sub-threshold analog circuits to digital filter banks Sub-threshold CMOS analog implementations have proven difficult because of the poor control over the gate thresholds, and hence the scatter in the center frequencies of the cochlear filters. Integrated digital filter banks are quite feasible, although the steep roll-offs of the cochlear filters and other time-domain cochlear stages have made the circuits excessively complex and unstable despite many simplifications.

The technology that we settled on is the switched capacitor filter (SCF). This was arrived at after an exhaustive look at passive, active RC, digital, and continuous time filter approaches.

Each of these had its advantages and disadvantages, and SCF's proved the best compromise. Details of our proposed approaches and preliminary fabricated circuits are discussed in the second part of this proposal.

II. Early Auditory Processing: Binaural Processing and Phonemic Segmentation

II.1 Early binaural processing

In binaural sound processing, the central auditory system compares the signals impinging on the two ears, detecting and utilizing various imbalances (e.g., sound level, time of arrival, and phase) to perform such perceptual tasks as sound localization in space and signal-to-noise enhancement. In this sense, binaural hearing is analogous to binocular vision in endowing perception with an extra spatial dimension based primarily on disparity measures in the stimulus projection upon the sensory organs. Numerous computational models have been proposed to account for these phenomena in vision and audition. In vision, most stereopsis algorithms detect and process spatial disparities between coincident images from the two retinae. In binaural models, instead, interaural disparities (such as phase and time delays) between signals from the two cochlei are usually derived from the phase-locked responses of the auditory-nerve through explicit time-domain operations.

An important example of the latter auditory algorithms is the Jeffress model. It postulates the existence of an organized array of *neural* delays to facilitate the computation of cross-correlation measures between the ipsilateral and contralateral cochlear outputs (Fig.4a). In such a model, an interaural-time delay (due, for instance, to a lateralized low frequency sound source) is effectively detected by a systematic comparison of the response pattern from one cochlea at a given instant, the response patterns of the other cochlea at various time-lags. The organized series of delays, therefore, provides both for copies of earlier cochlear outputs, and a spatial axis to encode and interpret the delays at which the maximum pattern match (or correlation) is achieved. The success of such correlation-based models in accounting for many psychophysical observations, and the convenience of their mathematical formulation, have indirectly lent support to, and acceptance of the notion of organized neural delay lines despite the lack of firm physiological evidence of their existence.

We have proposed a fundamentally different approach to the computation of interaural-differences. It emerges from an examination of the *spatial*, rather than temporal, disparities between the simultaneous travelling waves of the two ears (Fig.4b-c). Thus, a low frequency tone produces in each cochlea a spatially distributed travelling wave which is projected relatively intact unto the responses of the spatially ordered array of auditory-nerve-fibers. At any instant in time, the central binaural processor receives two spatial *images* (or snap-shots) of the travelling waves, one from each ear (Fig.4c). When the tone is centered, the images are identical; For binaurally unequal signals, however, the travelling waves differ systematically. Thus, when the tone is phase-shifted (or delayed) in one ear relative to the other, the images appear correspondingly shifted. Since this *spatial* disparity between the travelling waves is proportional to the *temporal* delays between the two ears, the binaural processing of all interaural-time-differences can be reduced to purely spatial operations. For instance, the network of "coincidence" detectors in Fig.4b performs these computations by effectively correlating the instantaneous images from the two ears at various relative horizontal shifts. The location of the maximal activity (correlation) in the plane of the network, and the sharpness of this profile, can be directly related to such psychophysical attributes as the lateralization of the

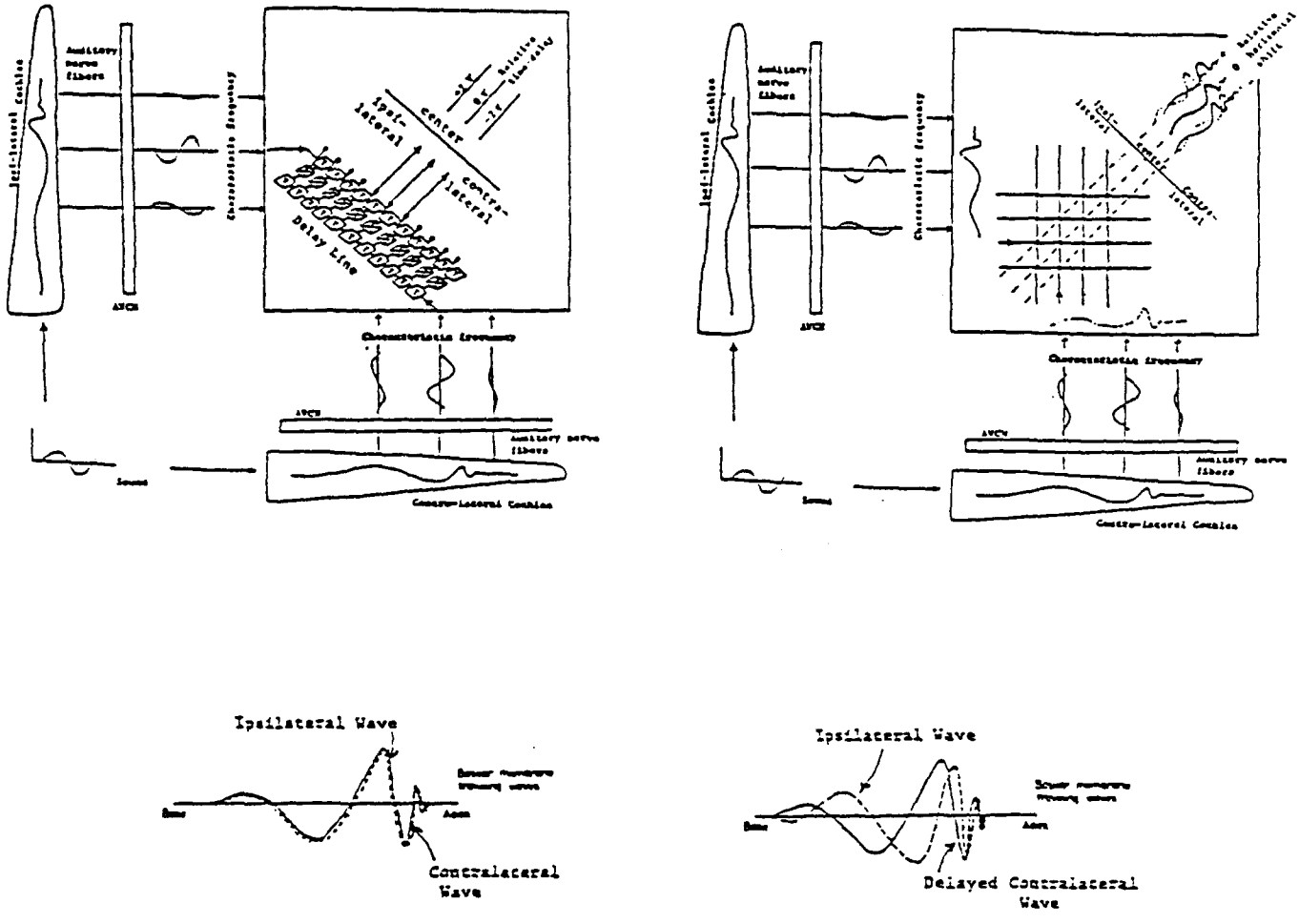


Figure 4: A schematic of early binaural auditory processing.

(a) The binaural processor proposed by Jeffress for the detection and encoding of interaural-time-differences. The input is the responses of the auditory-nerve-fiber array, conveyed in a tonotopically ordered manner via the pathways of the anteroventral nucleus (AVCN). At each node of the two dimensional binaural network, the responses of an ipsi-lateral fiber and a contra-lateral fiber *of an equal CF* are correlated at various relative delays. Depending on the initial interaural-time or phase-delay between the inputs, the correlation will be maximal at a specific location along the chain of *neural* delays. Therefore, along each diagonal array of nodes, the network effectively compares the travelling wave pattern from the ipsilateral ear with the *relatively time-delayed* pattern from the other ear. (b) The alternative *spatial* binaural processor. No functional neural time-delays are assumed. and the network correlates the responses of its ipsi-lateral and contra-lateral inputs in a matrix of nodes. Therefore, the inputs to each node are generally of unequal *CF's* (except along the major diagonal). Along each diagonal array, the network effectively compares the travelling wave pattern from the ipsi-lateral ear with the simultaneous, but *relatively spatially-shifted* pattern from the other ear. Since interaural-differences create proportional spatial disparities between the simultaneous patterns of activity from the two ears, the correlation will be maximal at a specific location (or spatial shift) depending on the original interaural-difference. (c) An illustration of the *horizontal and vertical* disparities between the simultaneous travelling wave patterns of the two ears that result from interaural-time-delays.

sound source and the degree of its compactness in space. Many other possible inequalities in binaural inputs, for instance in their envelopes, degree of correlation, or bandwidths, can be readily detected and consistently represented via the spatial disparities between the resulting travelling waves.

This network has recently been implemented both in hardware and fast software for real-time applications in speech recognition and pitch extraction. The hardware implementation came in the form of sub-threshold analog chip fabricated in Caltech by Dr. Carver Mead. The Software implementation on the Cray was carried out by Dr. Malcom Slany at Apple. Future experimentation with these implementations will be coordinated with the two above mentioned groups.

II.2 Phonemic segmentation: Unsupervised learning algorithms

The goal of this project is to develop *unsupervised* learning algorithms to configure networks able to segment the sound stream based on either pitch or timber cues. The motivation behind this work is two-fold: (1) From a theoretical standpoint, such networks can explain how such functions emerge spontaneously in the auditory nervous system; and (2) they can in turn be applied as pitch or phoneme detector devices. Being unsupervised, the network connectivities reflect the structure of the input set, and not that of an imposed external "teacher" as is, for instance, the case in the backpropogation learning algorithm.

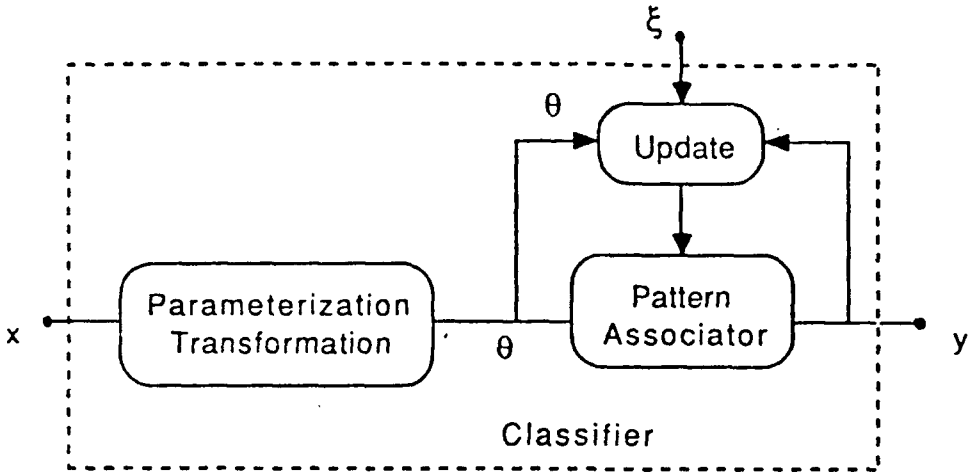
Such an adaptive pattern classifying algorithm was developed by T. Teolis (MS Thesis) (Fig.5). An overall block diagram of the adaptive algorithm is shon in Fig.5a. It is based on a choice of a similarity function which is restricted to operate only on the network outputs (Fig.5b). The learning rule is a gradient decent that aims to minimize an appropriate energy function ($g_\alpha(\cdot)$ with respect to an initial point that reflects the statistics of the input training set. The algorithm is computationally efficient and requires little storage capacity since it incorporates first order estimators of the input statistics rather than the actual input pattern sequence to compute its updates. The adaptive network has been demonstrated successfully as a segmentation algorithm of natural speech (Fig.5c) and of sequences of musical notes from various instruments. Details of this work are available in the report by T. Tiolis appended to this proposal.

III. Neurophysiology, Psychoacoustics, and Models of the Auditory Cortex

There are several ongoing projects to explore the functional organization of the primary auditory cortex. They range from neurophysiological mappings of the responses of the various areas of the auditory cortex, to psychoacoustical testing of the functional hypotheses, to the formulation of mathematical models of these data.

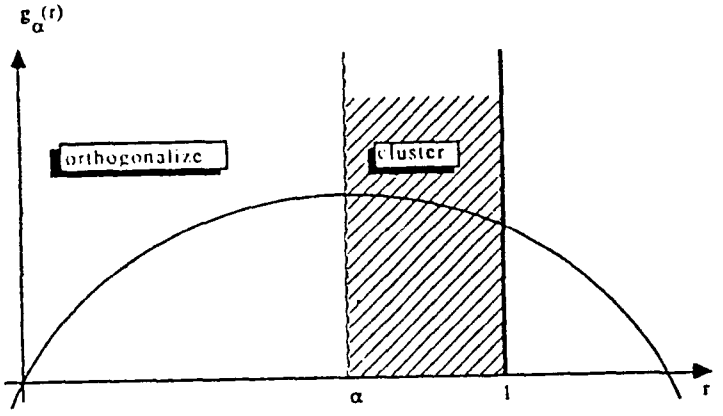
III.1 Neurophysiological mappings of the primary auditory cortex

The primary auditory cortex (AI) is essential for the perception and localization of sound. Its precise role in carrying out these functions, however, remains a mystery despite extensive knowledge gained from ablation experiments and from single and multi-unit recordings with various complex stimuli. At present, only two general organizational features of AI are firmly established: the spatially ordered tonotopic axis, and the alternating bands of binaural response properties that run perpendicularly to the isofrequency planes (Fig.6). They are roughly analogous to the retinotopic maps and the ocular dominance columns of the primary visual cortex.



Classification system model

(a)



The nonlinear function $g_\alpha(\cdot)$

(b)

(c)

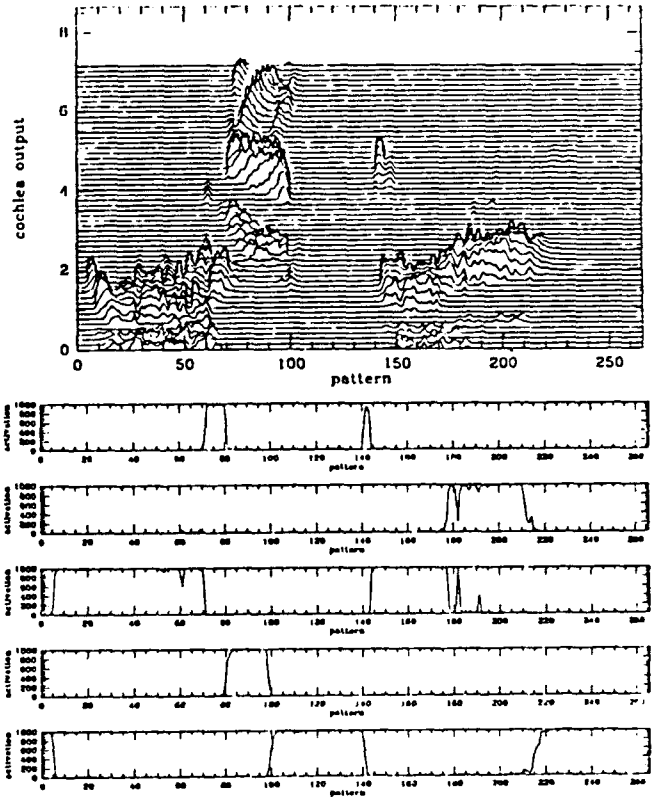


Figure 5: Unsupervised Classification and Segmentation. (a) The overall plan of the adaptive segmentation algorithm. (b) The nonlinear function used to cluster and classify different phoneme classes. (c) Top is the auditory output of the word "Six". The bottom traces are those of the segmentation algorithm. Associated with each box is a phoneme or output pattern that is flagged when the auditory output matches it. Thus there are five distinct patterns in this auditory stream, and each takes the indicated intervals of time.

While these axes are fundamental to the organization of AI, they only relate to basic simple properties of the acoustic stimulus that are already established at much lower levels of the auditory pathway. Tonotopic order, for instance, originates at the cochlea, while binaural columns exist at least as early as the inferior colliculus. With the exception of the more specialized auditory system of the bat, ordered responses to more complex stimulus features, analogous to the orientation columns and direction of motion selectivity in the visual cortex, have been more difficult to find in AI. At present, only a few reports hint at the existence of such maps in AI.

This issue was addressed directly in a series of experiments in the ferret AI. The study explored the detailed organization of the excitatory and inhibitory responses of cortical cells, i.e., their so-called receptive fields. The aim was to establish whether any systematic changes in the balance of inhibitory and excitatory responses occur in cells along the isofrequency planes and, if so, to determine the implications of these changes to responses to frequency-modulated (FM) tones and spectrally shaped noise stimuli. These response features are more complex than the determination of a single best frequency BF (tonotopicity) or the (binary) nature of a binaural interaction (e.g., an Excitatory-Excitatory or Excitatory-Inhibitory response). The receptive field of a cell represents, to first order, its transfer function, i.e., the way it filters or processes the input spectrum. Similarly, FM tones reveal information about the dynamic interplay between of the inhibitory and excitatory responses of the cell.

The basic findings of the above experiments can be summarized as follows (Fig.7):

- There seems to exist a spatially ordered change in the symmetry of the receptive fields in any given isofrequency plane in AI (Fig.7A). The receptive fields were measured in a series of mapping experiments. Briefly, cells sampled along the isofrequency contours were driven by a two-tone stimulus in which the first tone, $T1$, spanned a wide range of frequencies about the BF (± 1 or ± 2 octaves). The second tone, $T2$, was fixed at BF to provide a level of response against which the inhibition can be detected (Fig.8). $T1$ usually preceded $T2$ by a small interval (30 ms) so that the responses to each tone can be visually segregated in the rasters. An example of the responses of a cell with narrow and symmetric lateral inhibition is shown in the middle raster of Fig.8. The symmetry of the inhibition can also be seen in the combined spike count curve computed in the window 100–180 ms and illustrated below the raster. This response curve as a function of frequency is what is called here the *receptive field* of the cell. Two other types of responses were observed in cortical cells in which the receptive fields were significantly asymmetric. In one, inhibition was strong only from frequencies above the BF; in the other, the inhibition was largely from below the BF.

  Considering the results of experiments from over 20 animals, the outline of the distribution of three above response types can be described as follows (Fig.9): At the center of AI, units respond with a narrow excitatory tuning curve at BF, flanked by narrow symmetric inhibitory side-bands. The receptive fields become more asymmetric away from the center. In one direction (caudally in the ferret AI), the inhibitory side-bands above the BF become relatively stronger. The opposite occurs in the other direction. These response types tend to organize along one or more bands that parallel the tonotopic axis (i.e., orthogonal to the isofrequency planes).

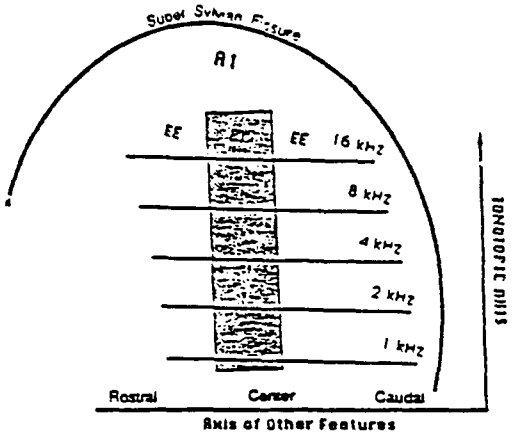


Figure 6: Schematic of ferret primary auditory cortex (AI).

The tonotopic axis runs in the mediolateral direction with low frequencies laterally. Isofrequency planes extend along the rostro-caudal axis. Presumed binaural columns intersect the isofrequency planes. Dimensions of AI vary considerably across animals, but average distance between octave frequencies is 0.5-1 mm.

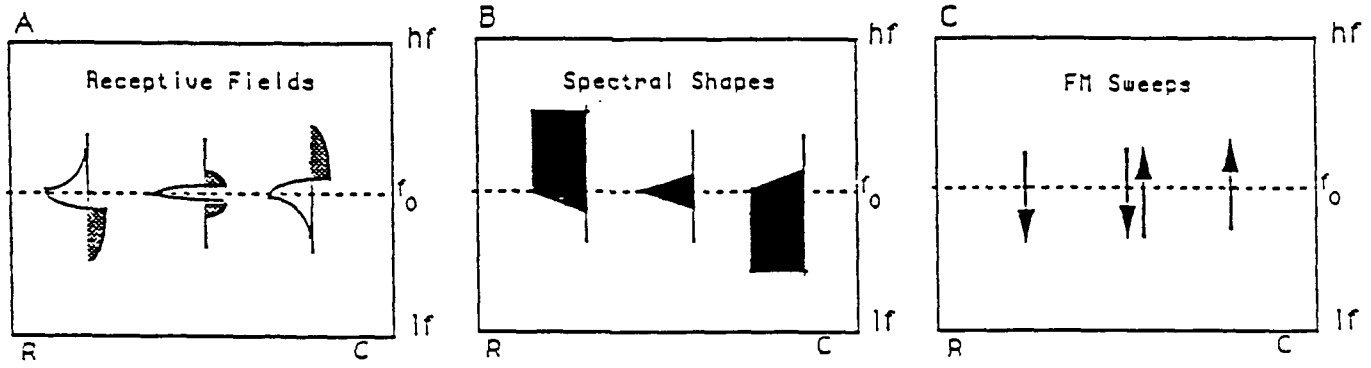


Figure 7: Schematic of basic findings of physiological experiments.

Schematic of basic findings of physiological experiments. Each box represents the AI, with ordinate representing the tonotopic axis from low (lf) to high (hf) frequencies, and the abscissa representing the (*R*)ostro-(*C*)audal axis. Dotted line represents a particular isofrequency plane at best frequency (*BF*) f_o. A: Schematic of receptive field organization in ferret AI. Each small plot represents the receptive field of a cell located along the indicated isofrequency plane, i.e., with $BF = f_o$. The excitatory and inhibitory responses as a function of frequency are symbolized by the clear and shaded regions, respectively. Cells near the center of AI have symmetric receptive fields, with narrow inhibition flanking a narrow excitatory tuning curve. Towards the edges, receptive fields become more asymmetric and broader. Inhibition from high frequencies becomes relatively predominant caudally and weak rostrally, while low frequency inhibition exhibits the opposite trend. B: Schematic of AI responses to spectrally shaped noise stimuli. Each small plot represents the spectrum of the most effective stimulus for cells at that location along the isofrequency plane. Near the center of AI, cells best respond to narrow-band noise centered around f_o. In the caudal region, cells respond best to stimuli that extend to lower than *BF* frequencies, and lack energy above the *BF*. The opposite is true for rostrally located cells. C: Schematic of AI responses to frequency-modulated (FM) tones. Each arrow represents the most effective direction of the FM sweep for cells at that location along the isofrequency plane. Near the center, cells are least selective, responding well to both directions. Caudally, cells become more responsive to sweeps from low to high frequencies, i.e. arriving from the receptive field region with least inhibition. The opposite occurs in the rostral region.

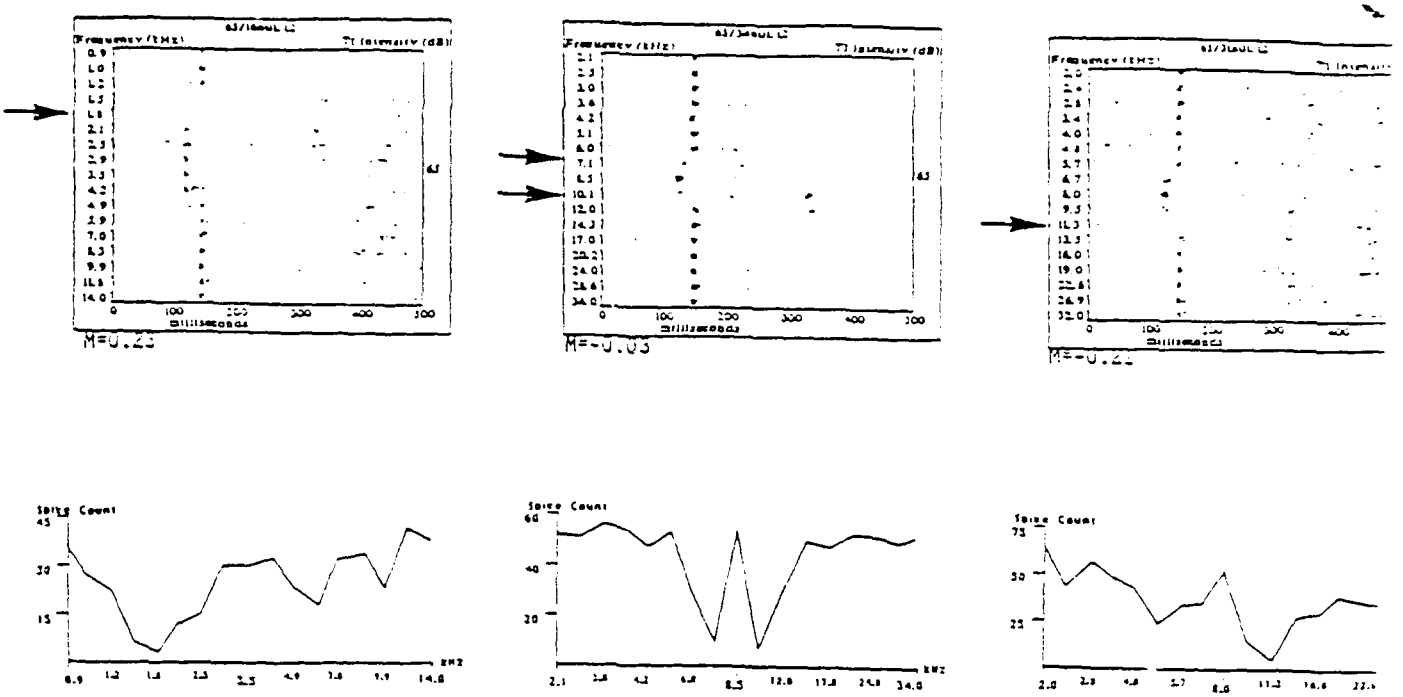


Figure 8: Typical responses of cells with three types of side-band inhibition.

Middle Raster and Plot: Raster of responses to a *two-tone* stimulus. The cell's *BF* is 8.5 kHz. *T*1 starts at 100 ms into the sweep; its intensity is indicated to the right of the raster (65 dB SPL); its frequency spans ±2 octaves around the *BF* in 1/4 octave steps. The second tone, *T*2, is delayed by 30 ms relative to *T*1; frequency = *BF*; intensity = 60 dB SPL. Ten repetitions are made at each *T*1 frequency. The phasic responses of the cell to the two tones are segregated because of the inter-tone delay. *T*1 responses start approximately 16 ms following the onset of the tone (i.e., at 116 ms), and are restricted to a narrow range of frequencies around the BF (8 kHz). *T*2 responses start at 144 ms. They are vigorous when *T*1 is not near *BF* and are suppressed otherwise. Lateral inhibition is evidenced by the near absence of responses to either tone at the frequencies marked by the two arrows. *Plot below the raster* shows the total spike count as a function of *T*1 frequency; the window is 80 ms long (100–180 ms). Left Raster and Plot Typical responses of a cell with asymmetric inhibition from below the *BF*. The arrow marks the frequency of the one-sided inhibition. *Plot below the raster* shows the corresponding total spike count curve. Right Raster and Plot Typical responses of a cell with asymmetric inhibition from above the *BF*. The arrow marks the frequency of the one-sided inhibition. *Plot below the raster* shows the corresponding total spike count curve.

- Cell responses to spectrally shaped noise are consistent with the symmetry of their receptive fields. For instance, cells with strong inhibition from above the BF are most responsive to stimuli that contain least spectral energy above the BF, i.e., stimuli with the opposite asymmetry. Since receptive field symmetry is ordered along the AI, then so is the local symmetry about the BF of the spectral envelope of the most effective stimulus (Fig.7B).

- The selectivity of a cell's response to the direction of an FM tone correlates strongly with the symmetry of its receptive fields. Specifically, cells with strong inhibition from frequencies above (below) the BF prefer upward (downward) moving sweeps (Fig.7C). Thus, selectivity to FM direction is also mapped along the isofrequency planes of the AI.

What are the basic attributes of the stimulus that are mapped in such representations? One such attribute that is immediately apparent from the experimental data is the direction of an FM sweep. It is conceivable that the observed receptive field organization solely functions to generate the FM maps. However, it is also possible that other stimulus features are extracted and mapped simultaneously, as is the case for instance in the primary visual cortex where selectivity to the direction of edge motion (analogous to FM) and orientation are functionally linked.

The second attribute of the acoustic spectrum that AI responses likely encode by their differential distribution along the isofrequency planes, is a local measure of the shape of the acoustic spectrum - specifically, the locally averaged *gradient* of the spectrum. This conjecture follows from the schematics of Fig.7B where best responses to spectral peaks or edges of different symmetries are mapped systematically across the AI. The significance of such a map stems from its enhancement and explicit representation of such perceptually important features as the shape of spectral peaks, edges, and the spectral envelope. This gradient map can be viewed as a one dimensional analogue of the orientation columns of the visual cortex, since the orientation of a two-dimensional edge simply entails specifying its gradients in two directions.

The similarity of cortical auditory and visual principles of processing is consistent with conclusions of studies into the generation of the neocortex and its subsequent parcellation into distinct areas. Particularly relevant for our arguments are experiments in which visual inputs from the optic nerve are induced in newborn ferrets to project to the auditory cortex through the medial geniculate body. In the adult brains of such animals, AI cells can be shown to possess many of the response characteristics typical of the normal primary visual cortex, such as orientation selectivity. These and other manipulations, such as the transplanting of pieces of fetal neocortex to different positions, have all pointed to the homogeneity of the neocortex at its early stages of development, and the importance of subsequent influences, especially through afferent inputs, in differentiating the adult neocortical areas.

III.2 Psychoacoustical studies of spectral shape perception

The experimental results described above suggested that specific features of the shape of the acoustic spectrum are being extracted and mapped in the cortex. If so, then it is likely that important consequences must exist regarding the perception of such spectra. Very little direct evaluation of such features as the sensitivity of subjects to the symmetry of spectral peaks and local gradients exist. So we have developed experimental set-ups and paradigms with the help

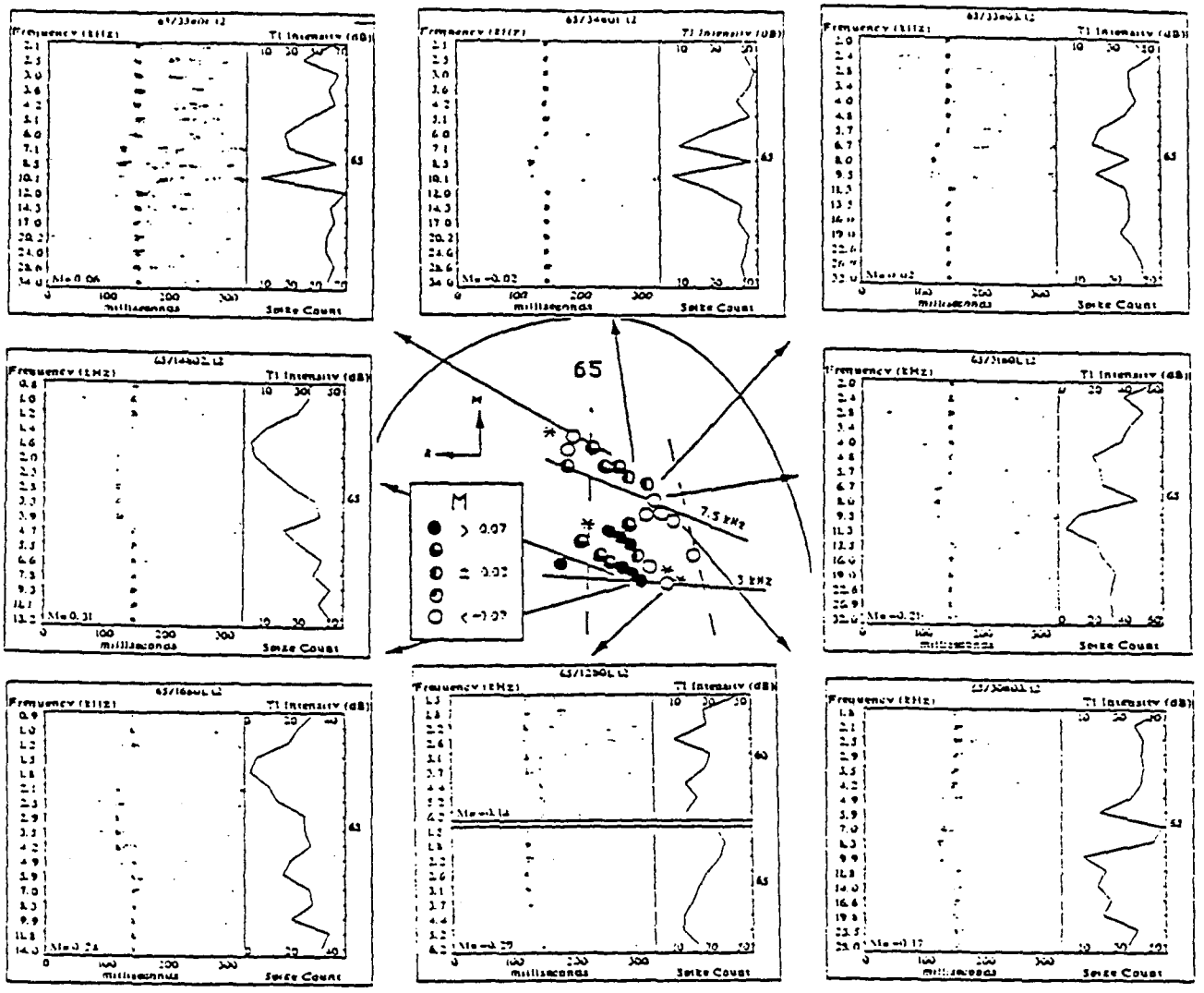


Figure 9: An example of the distribution of the *two-tone* response symmetry measure, M, in the ferret primary auditory cortex. M is defined as: $M = \frac{R_{>BF} - R_{<BF}}{R_{>BF} + R_{<BF}}$, where $R_{>BF}$ and $R_{<BF}$ are the total number of spikes (between 100-180 ms) for all frequencies above and below the BF, respectively. The circles indicate the locations of the electrode penetrations along the isofrequency contours (shown schematically as solid straight lines); asterisks mark penetrations with weak auditory responses. The arc represents the location of the supra-sylvian fissure; the dashed lines delineate the approximate borders of the band within which the M measure changes from extreme negative (clear circles) to extreme positive values (black circles). A key for the shading scheme used is shown on the left of the figure. The (M)edial and (R)ostral directions are indicated by the arrows; the arrow lengths represent 1/2 mm distances on the surface of the cortex. The rasters shown are those of cells sampled in the penetrations indicated by the arrows.

Marrian, C. R. K. and Peckerar, M. C. and Mack, I. and Pati, Y. C., "Electronic 'Neural' Nets for Solving Ill-posed problems", in *Maximum Entropy and Bayesian Methods*, 1989, 371–376, Kluwer Academic Publishers.

Pati, Yagyensh C. and Krishnaprasad, P. S., "Analysis and Synthesis of Feedforward Neural Networks Using Discrete Affine Wavelet Transforms", University of Maryland, Systems Research Center, Technical Report, SRC TR 90-44, 1990.

Teolis, Anthony and Pati, Yagyensh C. and Peckerar, Martin C. and Sham ma, Shihab", "Cascaded Neural-Analog Networks For Real Time Decomposition of Superposed Radar Return Signals in the Presence of Noise", University of Maryland, Systems Research Center, Technical Report SRC TR 89-33, 1989.

Personnel

The research reported here is documented in the theses and reports of the students (with degrees awarded) listed below. Several are now supported by the AFOSR grant towards their Ph.D. degree.
Y. Pati (MS), Geeth Chettiar (MS), James Fleshman (Res. Faculty, now at NIH), Philip Wiser (BS), Danniel Lin (MS), Tony Tiolis (MS), Preetham Gopalaswamy (MS), M. Niaming (MS), Svetlana Vranic (MS).